

第6章 关系数据理论



沈明玉

合肥工业大学

Hefei University of Technology 计算机与信息学院

第6章 关系数据理论

本章内容:

- 6.1 问题的提出
- 6.2 规范化
- 6.3 数据依赖的公理系统
- 6.4 关系模式的分解

合肥工业大学

Hefei University of Technology 计算机与信息学院

第6章 关系数据理论

6.1 问题的提出

- 回顾: 关系模式的形式化定义
 $R(U, D, DOM, F)$
- 什么是数据依赖?
 数据依赖是一种特殊的约束, 定义属性值间的相互关连(主要体现在值的相等与否)。
 - ✓ 一个关系内部属性与属性之间的约束关系;
 - ✓ 现实世界属性间相互联系的抽象;
 - ✓ 数据内在的性质;
 - ✓ 语义的体现。

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.1 问题的提出

- 数据依赖的类型:

函数依赖、多值依赖、连接依赖、包含依赖等。

- 认识函数依赖

在现实世界中, 每个学生的学号是唯一的(学校规定的), 因此当学号的值确定之后, 该学生的其他属性的值也就被确定了, 如该学生的姓名、性别等。-- **关系模式中的函数依赖**

- 数学中的函数: $Y = f(X)$
- 关系模式中函数依赖的表示: $X \rightarrow Y$

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.1 问题的提出

- 数据依赖对关系模式的影响

[例] 创建关系模式

$SDC(Sno, Sname, dno, dmanager, cname, grade)$

单一的关系模式: $SDC(U, F)$, U : 属性集 F : 函数依赖集

$U = \{Sno, Sname, dno, dmanager, cname, grade\}$

$F = \{Sno \rightarrow dno, Sno \rightarrow Sname, dno \rightarrow dmanager, (Sno, Cname) \rightarrow grade\}$

分析: 候选码?

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.1 问题的提出

分析表中数据:

Sno	sname	ssex	dno	dmanager	cno	grade
20160001	宋江	男	0001	蔡京	c211	NULL
20160001	宋江	男	0001	蔡京	c212	NULL
20160001	宋江	男	0001	蔡京	c213	NULL
20160001	宋江	男	0001	蔡京	c214	NULL
20160001	宋江	男	0001	蔡京	c215	NULL
20160001	宋江	男	0001	蔡京	c216	NULL
20160001	宋江	男	0001	蔡京	c217	NULL
20160001	宋江	男	0001	蔡京	c218	NULL
20160001	宋江	男	0001	蔡京	c219	NULL
20160002	卢俊义	男	0001	蔡京	c104	NULL
20160002	卢俊义	男	0001	蔡京	c208	NULL
20160002	卢俊义	男	0001	蔡京	c210	NULL
20160003	吴用	男	0001	蔡京	c104	NULL
20160003	吴用	男	0001	蔡京	c208	NULL
20160003	吴用	男	0001	蔡京	c210	NULL
20160004	公孙胜	男	0001	蔡京	c104	NULL
20160004	公孙胜	男	0001	蔡京	c208	NULL
20160004	公孙胜	男	0001	蔡京	c210	NULL
20160005	关胜	男	0001	蔡京	c104	NULL
20160005	关胜	男	0001	蔡京	c208	NULL

6.1 问题的提出

◆ 存在的问题

- **数据冗余**: dmanager
- **插入异常**: 该插入的数据无法插入。当某个学生尚未选课时, 该学生的基本信息无法插入。
- **删除异常**: 不该删除的数据被删除。若某个学生只选修了一门课程, 当删除该课程时, 该学生的基本信息也将一并删除。
- **修改异常**: 简单的修改问题操作过于复杂。当系主任变动时, 该系所有学生的dmanager属性值均需修改。

6.1 问题的提出

◆ 原因: 关系模式中存在“不好”的数据依赖!

◆ 解决方法: 通过分解关系模式来消除其中不合适的数据依赖。

如: 将单一的关系模式分成3个关系模式:

$S(Sno, Sname, dno) : Sno \rightarrow Sname, Sno \rightarrow dno;$

$S_C(Sno, Cname, Grade) : (Sno, Cname) \rightarrow Grade;$

$D(dno, dmanager) : dno \rightarrow dmanager;$

分解之后, 上述数据冗余和操作异常问题得到解决!

6.1 问题的提出

◆ 一个好的关系模式应该具备以下四个条件:

- 尽可能少的数据冗余;
- 没有插入异常;
- 没有删除异常;
- 没有更新异常。

第6章 关系数据理论

6.2 规范化

规范化理论是用来改造关系模式, 通过分解关系模式来消除其中不合适的数据依赖, 以解决插入异常、删除异常、更新异常和数据冗余问题。

6.2 规范化

6.2.1 函数依赖

定义6.1: 设 $R(U)$ 是属性集 U 上的关系模式, X 和 Y 是 U 的子集。若对于 $R(U)$ 的任一可能的关系 r , r 中不可能存在两个元组在 X 上的属性值相等, 而在 Y 上的属性值不等, 则称 **X 函数确定 Y** 或 **Y 函数依赖于 X** , 记作:

$$X \rightarrow Y \quad (X \text{ 为决定因素})$$

- 函数依赖属于语义范畴的概念;
- 设计者根据应用需求可以对现实世界作强制的规定;
- 所有关系实例均要满足的约束条件。

如: $Sno \rightarrow dno, Sno \rightarrow Sname, dno \rightarrow dmanager,$

$(Sno, Cname) \rightarrow grade$

6.2 规范化

◆ 需要区分的几种函数依赖:

- ✓ 平凡的函数依赖: $(Sno, sname) \rightarrow Sno$
- ✓ 非平凡的函数依赖: $Sno \rightarrow dno$
- ✓ 完全函数依赖: $(Sno, Cname) \rightarrow grade$
- ✓ 部分函数依赖: $(Sno, Cname) \rightarrow Sname$
- ✓ 传递函数依赖: 由 $Sno \rightarrow dno, dno \rightarrow dmanager$, 得到:
 $Sno \rightarrow dmanager$ (传递依赖)

6.2 规范化

6.2.2 码

定义6.2: 设 K 为 $R(U, F)$ 中的属性或属性组, 若 $K \xrightarrow{F} U$, 则 K 为 R 的候选码。

- 若 $K \xrightarrow{F} U$, 则 K 为超码。候选码是最小的超码, 即的任意一个真子集都不是候选码。
- 若 K 唯一, 则 K 是主码, 若有多个候选码, 则选择一个作为主码。
- 主属性: 包含在任一候选码中的属性都是主属性, 否则为非主属性。
- 全码: R 中所有属性构成唯一的候选码 (主码)。
- 外码: (回顾外码的定义)

6.2 规范化

6.2.3 范式

- ◆ 范式是符合某一种级别的关系模式的集合。
- ◆ 关系数据库中的关系必须满足一定的要求。满足不同程度要求的为不同范式。
- ◆ 范式的种类:
 - 第一范式(1NF)、第二范式(2NF)、第三范式(3NF)、BC范式(BCNF)、第四范式(4NF)等。

各范式之间的关系: $1NF \subset 2NF \subset 3NF \subset BCNF \subset 4NF$

合肥工业大学 Hefei University of Technology 计算机与信息学院

6.2 规范化

6.2.4 第一范式 (1NF) -- 关系模式的基本要求

定义: 如果一个关系模式 R 的所有属性都是不可分的基本数据项, 则 $R \in 1NF$ 。

分析: 关系模式 $sd(Sno, Sname, dno, dmanager, cname, grade) \in 1NF$

主码: $(Sno, Cname) \quad (Sno, Cname) \rightarrow dno$

存在: 非主属性 (dno) 对码 $(sno, cname)$ 的部分函数依赖!

6.2 规范化

6.2.5 第二范式 (2NF)

定义6.3: 若 $R \in 1NF$, 且每一个非主属性完全函数依赖于码, 则 $R \in 2NF$ 。

例: $sd(Sno, Sname, dno, dmanager, cname, grade) \notin 2NF$

分解后: $sc(Sno, Cname, Grade) \in 2NF$

$sd(Sno, sname, dno, dmanager) \in 2NF$

发现: 达到2NF的关系模式, 还可能存在数据冗余和异常问题。(sd)

- 没有学生信息, 系的信息无法插入;
- 删除学生信息, 可能导致系的信息丢失;
- 修改系主任的操作依然复杂 (冗余导致)。

合肥工业大学 Hefei University of Technology 计算机与信息学院

合肥工业大学 Hefei University of Technology 计算机与信息学院

6.2 规范化

原因: $sd(sno, sname, dno, dmanager)$ 中函数依赖:

$(sno \rightarrow dno, dno \rightarrow dmanager) \Rightarrow sno \rightarrow dmanager$

- ✓ SD 中存在非主属性 ($dmanager$) 对码的传递函数依赖!
- ✓ 这种函数依赖是导致冗余和异常的直接原因;
- ✓ 应设法消除关系模式中非主属性对码的传递函数依赖。

6.2 规范化

6.2.6 第三范式 (3NF)

定义6.4: 设关系模式 $R(U, F) \in 2NF$, 若 R 中所有的非主属性都直接依赖于码, 则称:

$R(U, F) \in 3NF$

- 若 $R \in 3NF$, 则每一个非主属性既不部分依赖于码也不传递依赖于码

上例中, $SD \notin 3NF$

- 不允许非主属性之间存在依赖关系!

合肥工业大学 Hefei University of Technology 计算机与信息学院

合肥工业大学 Hefei University of Technology 计算机与信息学院

6.2 规范化

◆ 问题的解决

采用投影分解法, 把 SD 分解为两个关系模式, 消除传递依赖:

$$S(sno, sname, dno)$$

$$d(dno, dmanager)$$

s 的码为 sno , d 的码为 dno .

- 分解后的关系模式 s 与 d 中不再存在传递依赖, 均属于 $3NF$.

Hefei University of Technology 计算机与信息学院

6.2 规范化

6.2.7 BC范式 (BCNF)

◆ 问题分析

关系模式 $STJ(S, T, J)$, S 表示学生, T 表示教师, J 表示课程。

应用语义:

- 每个教师只教一门课;
- 每门课可以有多个教师;
- 每个学生选定某门课, 就对应一个固定的教师。

函数依赖:

$(S, J) \rightarrow T, (S, T) \rightarrow J, T \rightarrow J$, 找出候选码、分析范式!

Hefei University of Technology 计算机与信息学院

6.2 规范化

学生	教师	课程
张三	赵刚	数据库
张三	刘平	网络技术
李四	李明	数据库
王五	赵刚	数据库
赵六	李勇	数据库
李四	张磊	网络技术
王五	刘平	网络技术
赵六	张磊	网络技术

问题:

数据冗余: 每个教师只教一门课, 但每个选修该教师所开课程的学生元组都要记录这一信息。

插入异常: 当某门没有学生选修, 因为主属性不能为空, 教师上该门课程的信息就无法插入。

删除异常: 删除学生选课记录的同时, 可能导致教师开设该门课程信息的删除。

更新异常: 当某个教师开设的课程改名后, 所有选修该教师该门课程的学生元组都要进行修改, 如果漏改某个数据, 则破坏了数据的完整性。

原因: 存在主属性(J) 对于一个不包含它的码 (S, T) 的部分或传递依赖!

Hefei University of Technology 计算机与信息学院

6.2 规范化

定义6.5: 关系模式 $R(U, F) \in 1NF$, 若 $X \rightarrow Y$ 且 $Y \not\subseteq X$ 时 X 必含有码, 则

$$R(U, F) \in BCNF.$$

- 等价于: 每一个决定属性因素都包含码。

- 若 $R \in BCNF$:

- ✓ 所有非主属性对每一个码都是完全函数依赖;
- ✓ 所有的主属性对每一个不包含它的码, 也是完全函数依赖;
- ✓ 没有任何属性完全函数依赖于非码的任何一组属性。

■ $R \in BCNF \iff R \in 3NF$

Hefei University of Technology 计算机与信息学院

6.2 规范化

- $STJ \in 3NF$

没有任何非主属性对码传递依赖或部分依赖!

- $STJ \notin BCNF$

T 是决定因素 ($T \rightarrow J$), T 不包含码。

- 解决方法: 将 STJ 分解为二个关系模式:

$$ST(S, T) \in BCNF, \quad TJ(T, J) \in BCNF$$

☆ 对函数依赖及其规范化的简单小结

Hefei University of Technology 计算机与信息学院

6.2 规范化

6.2.8 多值依赖 (MVD)

■ 例题分析

设有一学校关系模式: $R(C, T, B)$

其中: C - 课程, T - 教师, B - 参考书

应用语义:

- 一门课程由多个教师讲授;
- 同一门课程必须使用相同的一套参考书;
- 每个教师可以讲授多门课程;
- 每本参考书可以供多门课程使用。

Hefei University of Technology 计算机与信息学院

6.2 规范化

■ 得到如下非规范化关系:

课程 C	教员 T	参考书 B
物理	{李勇 王军}	{普通物理学 光学原理 物理习题集}
数学	{李勇 张平}	{数学分析 微分方程}
计算数学	{张平 周峰}	{高等代数 数学分析 ...}

Hefei University of Technology 计算机与信息学院

6.2 规范化

■ 用规范化的二维表表示:

课程C	教员T	参考书B
物理	李勇	普通物理学
物理	李勇	光学原理
物理	李勇	物理习题集
物理	王军	普通物理学
物理	王军	光学原理
物理	王军	物理习题集
数学	李勇	数学分析
数学	李勇	微分方程
数学	李勇	高等代数
数学	张平	数学分析
数学	张平	微分方程
数学	张平	高等代数

Hefei University of Technology 计算机与信息学院

6.2 规范化

◆ $R \in BCNF$

◆ R 具有唯一候选码 (C, T, B) , 即全码

◆ 关系模式 $R(C, T, B)$ 存在的问题:

- (1) 数据冗余度大
- (2) 插入操作复杂
- (3) 删除操作复杂
- (4) 修改操作复杂

存在
多值依赖!

Hefei University of Technology 计算机与信息学院

6.2 规范化

■ 多值依赖

定义6.6: 设 $R(U)$ 是一个属性集 U 上的一个关系模式, X, Y 和 Z 是 U 的子集, 并且 $Z = U - X - Y$ 。关系模式 $R(U)$ 中多值依赖 $X \twoheadrightarrow Y$ 成立, 当且仅当对 $R(U)$ 的任一关系 r , 给定的一对 (x, z) 值, 有一组 Y 的值, 这组值仅由 x 值决定, 而与 z 值无关。

根据定义发现, 关系模式 $R(C, T, B)$ 中存在多值依赖:

$$C \twoheadrightarrow T, C \twoheadrightarrow B$$

Hefei University of Technology 计算机与信息学院

6.2 规范化

■ 平凡多值依赖和非平凡的多值依赖

- ◆ 若 $X \twoheadrightarrow Y$, 而 $Z = \emptyset$, 则称 $X \twoheadrightarrow Y$ 为平凡的多值依赖;
- ◆ 否则称 $X \twoheadrightarrow Y$ 为非平凡的多值依赖。

■ 多值依赖的性质

- (1) 多值依赖具有对称性 若 $X \twoheadrightarrow Y$, 则 $Y \twoheadrightarrow X$, 其中 $Z = U - X - Y$
- (2) 多值依赖具有传递性 若 $X \twoheadrightarrow Y$, $Y \twoheadrightarrow Z$, 则 $X \twoheadrightarrow Z - Y$
- (3) 函数依赖是多值依赖的特殊情况 若 $X \rightarrow Y$, 则 $X \twoheadrightarrow Y$ 。
- (4) 若 $X \twoheadrightarrow Y$, $X \twoheadrightarrow Z$, 则 $X \twoheadrightarrow Y \cup Z$ 。
- (5) 若 $X \twoheadrightarrow Y$, $X \twoheadrightarrow Z$, 则 $X \twoheadrightarrow Y \cap Z$ 。
- (6) 若 $X \twoheadrightarrow Y$, $X \twoheadrightarrow Z$, 则 $X \twoheadrightarrow Y - Z$, $X \twoheadrightarrow Z - Y$ 。

Hefei University of Technology 计算机与信息学院

6.2 规范化

■ 多值依赖与函数依赖的两个基本区别

- ◆ 多值依赖的有效性与其属性集的范围有关。若 $X \twoheadrightarrow Y$ 若在 U 上成立, 则在 $W(XY \subseteq W \subseteq U)$ 上一定成立; 反之则不然, 即 $X \twoheadrightarrow Y$ 若在 $W(W \subseteq U)$ 上成立, 在 U 上不一定成立。因为多值依赖不仅涉及 X, Y , 还与 $Z (Z = U - X - Y)$ 有关。(对比函数依赖!)
- ◆ 若函数依赖 $X \rightarrow Y$ 在 $R(U)$ 上成立, 则对于任何 $Y' \subset Y$ 均有 $X \rightarrow Y'$ 成立。而多值依赖 $X \twoheadrightarrow Y$ 若在 $R(U)$ 上成立, 却不能断言对于任何 $Y' \subset Y$ 有 $X \twoheadrightarrow Y'$ 成立。

Hefei University of Technology 计算机与信息学院

6.2 规范化

6.2.9 第四范式4NF

定义6.7: 关系模式 $R(U, F) \in 1NF$, 如果对于 R 的每个非平凡多值依赖

$X \twoheadrightarrow Y (Y \not\subseteq X)$, X 都含有码, 则 $R \in 4NF$ 。

- 如果 $R \in 4NF$, 则 $R \in BCNF$
- 不允许有非平凡且非函数依赖的多值依赖
- 允许的非平凡多值依赖必须存在对应的函数依赖

6.2 规范化

上例: $R(C, T, B) \notin 4NF$

因为: 存在非平凡的多值依赖 $C \twoheadrightarrow T$, 且 C 不是码。

- 用投影分解法把 R 分解为如下两个关系模式:

$CT(C, T) \in 4NF$

$CB(C, B) \in 4NF$

$C \twoheadrightarrow T, C \twoheadrightarrow B$ 是平凡多值依赖。

6.2 规范化

6.2.10 规范化小结

- 规范化的基本思想: 通过对关系模式的分解, 逐步消除不好的数据依赖, 以解决关系模式中存在的数据库冗余和操作异常问题。
- 规范化的本质: 概念的单一化。如关系模式

$sdc(Sno, Sname, dno, dmanager, cname, grade)$ 中描述了三个概念:

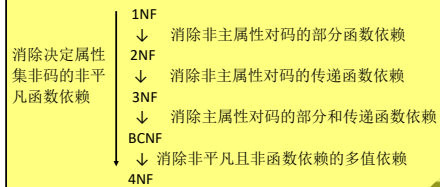
学生、系、学生选课。分解之后得到三个子模

$S(Sno, Sname); D(dno, dmanager); S_C(Sno, Cname, Grade);$

每个子模式均描述一个单一的概念, 即: 一个中心!

6.2 规范化

- 关系模式规范化的基本步骤



第6章 关系数据理论

6.3 数据依赖的公理系统

6.3.1 逻辑蕴涵

定义6.8: 设 F 是在关系模式 $R(U)$ 上成立的函数依赖集合, X, Y 是属性集 U 的子集, $X \rightarrow Y$ 是一个函数依赖。如果从 F 中能够推导出 $X \rightarrow Y$, 即如果对于 R 的每个满足 F 的关系 r 也满足 $X \rightarrow Y$, 则称 $X \rightarrow Y$ 为 F 的**逻辑蕴涵** (或 F 逻辑蕴涵 $X \rightarrow Y$), 记为 $F \models X \rightarrow Y$ 。

- 码的求解与函数依赖有关
- 如何从函数依赖集 F 中求得被 F 蕴涵的函数依赖
- 需要一套推理规则: Armstrong推理规则

6.3 数据依赖的公理系统

6.3.2 Armstrong公理系统

对于关系模式 $R(U, F)$, 有以下的推理规则:

- A1. **自反律**: 若 $Y \subseteq X \subseteq U$, 则 $X \rightarrow Y$ 为 F 所蕴涵。
- A2. **增广律**: 若 $X \rightarrow Y$ 为 F 所蕴涵, 且 $Z \subseteq U$, 则 $XZ \rightarrow YZ$ 为 F 所蕴涵。
- A3. **传递律**: 若 $X \rightarrow Y$ 及 $Y \rightarrow Z$ 为 F 所蕴涵, 则 $X \rightarrow Z$ 为 F 所蕴涵。

定理6.1: Armstrong推理规则是正确的。

(推理规则请同学们参照教材内容理解!)

6.3 数据依赖的公理系统

◆ 根据A1, A2, A3这三条推理规则可以得到下面三条推理规则:

- 合并规则: 由 $X \rightarrow Y, X \rightarrow Z$, 有 $X \rightarrow YZ$. (A2, A3)
- 伪传递规则: 由 $X \rightarrow Y, WY \rightarrow Z$, 有 $XW \rightarrow Z$. (A2, A3)
- 分解规则: 由 $X \rightarrow Y$ 及 $Z \subseteq Y$, 有 $X \rightarrow Z$. (A1, A3)

◆ 根据合并规则和分解规则, 可得:

引理6.1: $X \rightarrow A_1 A_2 \dots A_k$ 成立的充分必要条件是 $X \rightarrow A_i$ 成立 ($i=1, 2, \dots, k$).

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.3 数据依赖的公理系统

6.3.3 函数依赖的闭包

定义6.9: 在关系模式 $R(U, F)$ 中为 F 所逻辑蕴含的函数依赖的全体叫作 F 的闭包, 记为 F^+ . (函数依赖集 F 的闭包)

● Armstrong公理系统是有效的、完备的

- 有效性: 由 F 出发根据 Armstrong 公理推导出来的每一个函数依赖一定在 F^+ 中;
- 完备性: F^+ 中的每一个函数依赖, 必定可以由 F 出发根据 Armstrong 公理推导出来。

定义6.10: 设 F 为属性集 U 上的一组函数依赖, $X \subseteq U$,

$X_F^+ = \{A \mid X \rightarrow A \text{ 能由 } F \text{ 根据 Armstrong 公理导出}\}$, X_F^+ 称为属性集 X 关于函数依赖集 F 的闭包. (属性集 X 的闭包)

6.3 数据依赖的公理系统

引理6.2: 设 F 为属性集 U 上的一组函数依赖, $X, Y \subseteq U, X \rightarrow Y$ 能由 F 根据 Armstrong 公理导出的充分必要条件是 $Y \subseteq X_F^+$.

◆ 引理6.2的用途: 将判定 $X \rightarrow Y$ 是否能由 F 根据 Armstrong 公理导出的问题, 转化为求出 X_F^+ 、判定 Y 是否为 X_F^+ 的子集的问题。

[例] 已知关系模式 $R(U, F)$, 其中: $U = \{A, B, C, D, E\}$:

$F = \{AB \rightarrow C, B \rightarrow D, C \rightarrow E, EC \rightarrow B, AC \rightarrow B\}$.

求 $(AB)_F^+$.

解: 设 $X^{(0)} = AB$:

- (1) $X^{(1)} = AB \cup CD = ABCD$.
- (2) $X^{(2)} = X^{(1)} \cup E = ABCDE$.
- (3) $X^{(2)} = U$, 算法终止 $\Rightarrow (AB)_F^+ = ABCDE$. (AB 为候选码)

6.3 数据依赖的公理系统

定义6.11: 关系模式 $R(U)$ 的两个函数依赖集 F 和 G , 如果满足 $F^+ = G^+$, 则称 F 和 G 是等价的函数依赖集。记作: $F \equiv G$. 如果 F 和 G 等价, 就说 F 覆盖 G , 或 G 覆盖 F .

引理6.3: $F^+ = G^+$ 的充分必要条件是 $F^+ \subseteq G^+$ 和 $G^+ \subseteq F^+$.

定义6.12: 设 F 是属性集 U 上的函数依赖集。如果 F 满足下列四个条件, 则称 F 为极小函数依赖集 F_{\min} :

- (1) $F_{\min}^+ = F^+$;
- (2) 每个函数依赖的右边都是单属性;

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.3 数据依赖的公理系统

(3) F_{\min} 中没有冗余的函数依赖 (即在 F_{\min} 中不存在这样的函数依赖 $X \rightarrow Y$, 使得 F_{\min} 与 $F_{\min} - \{X \rightarrow Y\}$ 等价), 即减少任何一个函数依赖都将与原来的 F 不等价;

(4) 每个函数依赖的左边没有冗余的属性 (即 F_{\min} 中不存在这样的函数依赖 $X \rightarrow Y$, X 有真子集 W 使得 $F_{\min} - \{X \rightarrow Y\} \cup \{W \rightarrow Y\}$ 与 F_{\min} 等价), 减少任何一个函数依赖左部的属性后, 都将与原来的 F 不等价。

定理6.2: 每一个函数依赖集 F 均等价于一个极小函数依赖集 F_{\min} , 此 F_{\min} 称为 F 的最小依赖集。

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.3 数据依赖的公理系统

6.3.4 候选码的一种求解方法

■ 属性分类:

对于给定的关系模式 $R(A_1, \dots, A_n)$ 和函数依赖集 F , 可将其属性分为以下四类:

L 类属性: 只在函数依赖左边出现的属性

R 类属性: 只在函数依赖右边出现的属性

N 类属性: 函数依赖的左右两边均未出现过的属性

LR 类属性: 函数依赖的两边都出现过的属性

合肥工业大学

Hefei University of Technology 计算机与信息学院

6.3 数据依赖的公理系统

■ 求解思路:

- (1) 若 $X(X \in R)$ 是 L 类属性, 则 X 必为 R 的某一候选码的成员。
- (2) 若 $X(X \in R)$ 是 L 类属性, 且 X_F^+ 包含了 R 的全部属性, 则 X 必为 R 的唯一候选码。
- (3) 若 $X(X \in R)$ 是 R 类属性, 则 X 不在任何候选码中。
- (4) 若 $X(X \in R)$ 是 N 类属性, 则 X 包含在 R 的某一候选码中。
- (5) 若 $X(X \in R)$ 是 R 的 N 类和 L 类属性组成的属性集, 且 X_F^+ 包含了 R 的全部属性, 则 X 是 R 的唯一候选码。

Hefei University of Technology 计算机与信息学院

6.3 数据依赖的公理系统

■ 多属性函数依赖集候选码的求解算法

- (1) 属性分类 (L 、 R 、 N 和 LR)。设: X 中包含 L 类和 N 类的所有属性, Y 中包含 LR 类的所有属性。
- (2) 若 X_F^+ 包含了 R 的全部属性, 则 X 为唯一的候选码, 转 (5); 否则, 转 (3)。
- (3) 在 Y 中取一个属性 A , 求 $(XA)_F^+$, 若它包含了 R 的全部属性, 则转 (4); 否则, 调换一属性反复进行这一过程, 直到试完所有 Y 中的属性。
- (4) 如果已找出所有候选键, 则转 (5); 否则在 Y 中依次取两个、三个、..., 求它们的属性集的闭包, 直到其闭包包含 R 的全部属性。
- (5) 停止, 输出结果。

Hefei University of Technology 计算机与信息学院

第6章 关系数据理论

6.4 关系模式的分解

6.4.1 概述

定义6.13: 设有关系模式 $R(U)$, $R = R_1 \cup R_2 \cup \dots \cup R_n$, $\rho = \{R_1, R_2, \dots, R_n\}$ 。这里 ρ 称为 R 的一个分解, 也称为数据库模式。

■ 分解的目的

使模式更加规范化, 减少乃至消除数据冗余和更新异常。

■ 分解的三种准则

- ✓ 分解具有无损连接性
- ✓ 分解要保持函数依赖
- ✓ 分解既要具有无损连接性, 又要保持函数依赖

Hefei University of Technology 计算机与信息学院

6.4 关系模式的分解

6.4.2 具有无损连接的分解

定义6.14: 关系模式 $R(U, F)$ 的一个分解

$$\rho = \{R_1(U_1, F_1), R_2(U_2, F_2), \dots, R_n(U_n, F_n)\}$$

若 R 与 R_1, R_2, \dots, R_n 自然连接的结果相等, 则称关系模式 R 的这个分解 ρ 具有无损连接性。

- ◆ 反映了模式分解的**数据等价**原则, 保证不丢失信息。
- ◆ 不一定能解决操作异常和数据冗余等问题。

Hefei University of Technology 计算机与信息学院

6.4 关系模式的分解

● 一种判断分解是否具有无损连接性的方法

若关系模式 $R(U, F)$ 分解为两个子模式 $R_1(U_1, F_1)$ 和 $R_2(U_2, F_2)$, 该分解具有无损连接性的充分必要条件是:

- (1) $(R_1 \cap R_2) \rightarrow (R_1 - R_2) \in F^+$
或
- (2) $(R_1 \cap R_2) \rightarrow (R_2 - R_1) \in F^+$

Hefei University of Technology 计算机与信息学院

6.4 关系模式的分解

6.4.3 保持函数依赖的分解

定义6.14: 设关系模式 $R(U, F)$ 被分解为若干个关系模式:

$$\rho = \{R_1(U_1, F_1), R_2(U_2, F_2), \dots, R_n(U_n, F_n)\}$$

(其中 $U = U_1 \cup U_2 \cup \dots \cup U_n$, 且不存在 $U_i \subseteq U_j$, F_i 为 F 在 U_i 上的投影), 若 F 所逻辑蕴含的函数依赖一定也由分解得到的某个关系模式中的函数依赖 F_i 所逻辑蕴含, 则称关系模式 R 的这个分解 ρ 是保持函数依赖的。

- 保持函数依赖反映了分解的**依赖等价**原则。
- 依赖等价保证了分解后的模式与原有的模式在语义上的一致性。

Hefei University of Technology 计算机与信息学院

6.4 关系模式的分解

- ◆ 对关系模式的分解，使之属于第2、第3范式，只要采用规范的方法，既能实现无损连接，又能保持函数依赖。
- ◆ 若要分解后的模式属于BC范式，即使采用规范的方法，也只能保证无损连接，而不能绝对保持依赖。
- ◆ 保留适当冗余，达到以空间换时间的目的，也是模式分解的一个重要原则。

合肥工业大学

Hefei University of Technology 计算机与信息学院



6.4 关系模式的分解

6.4.4 模式分解的两个规则（分解为第2、第3范式）

1) 公共属性共享

保留公共属性，进行自然连接是分解后的模式实现无损连接的必要条件。

$sdc(Sno, sname, dno, dmanager, cname, grade)$

- 若存在部分依赖，决定因素的真子集 (Sno) 为公共属性。将模式分解成部分依赖和完全依赖两个子模式：

$sd(Sno, sname, dno, dmanager), sc(Sno, cname, grade)$

- 若存在传递依赖，则传递链的中间属性 (dno) 为公共属性，将链断开，分解成两个模式（保持了两个基本依赖）：

$s(Sno, sname, dno), d(dno, dmanager)$

合肥工业大学

Hefei University of Technology 计算机与信息学院



6.4 关系模式的分解

2) 相关属性合一

- 将以函数依赖形式联系在一起的相关属性放在一个模式中，从而使原有的函数依赖得以保持。这是分解后的模式实现保持依赖的充分条件。
- 对于存在部分依赖或传递依赖的相关属性则不应放在一个模式中，因为这正是导致数据冗余和更新异常的根源，也正是模式分解所要解决的问题。

合肥工业大学

Hefei University of Technology 计算机与信息学院



6.4 关系模式的分解

6.4.5 分解为BC范式的方法

设关系模式为 $R(A, B, C)$ ，其中 A, B, C 均为属性集，若存在违背BC范式的函数依赖 $A \rightarrow B$ (如 A 不包含码)，则以BC范式的违例 ($A \rightarrow B$) 为基础把关系模式分解为：

(1) $\{A, B\}$

(2) $\{A, C\}$ 或 $\{U - B\}$

如： $STJ(S, T, J) : (S, J) \rightarrow T, (S, T) \rightarrow J, T \rightarrow J$

分解： $TJ(T, J), TS(T, S)$

分析：无损连接、非保持依赖！

合肥工业大学

Hefei University of Technology 计算机与信息学院



6.4 关系模式的分解

6.4.6 分解为第4范式的方法

设关系模式为 $R(A, B, C)$ ，其中 A, B, C 均为属性集，若存在违背第4范式的非平凡多值依赖 $A \twoheadrightarrow B$ (如 A 不包含码)，则以第4范式的违例 ($A \twoheadrightarrow B$) 为基础把关系模式分解为：

(1) $\{A, B\}$

(2) $\{A, C\}$ 或 $\{U - B\}$

如： $R(C, T, B) C \twoheadrightarrow T, C \twoheadrightarrow B$

分解： $CT(C, T) \in 4NF, CB(C, B) \in 4NF$

合肥工业大学

Hefei University of Technology 计算机与信息学院



第6章 关系数据理论

■ 本章思考题

关系数据理论的实际指导意义？

■ 本章作业

1. 如何理解数据库的规范化设计？

2. P202 第2题

3. P203 第6题

合肥工业大学

Hefei University of Technology 计算机与信息学院

